

DATA NAME: NLCD U.S.G.S. National Land Cover Characterization

GRID NAME: d<xxxxx>

GRID AREA: 1x1 Degree block tiles

LOCATION: GIS Raster:

Introduction

This land cover data set was produced as part of a cooperative project between the U.S. Geological Survey (USGS) and the U.S. Environmental Protection Agency (USEPA) to produce a consistent, land cover data layer for the conterminous U.S. based on 30-meter Landsat thematic mapper (TM) data. National Land Cover Data (NLCD) was developed from TM data acquired by the Multi-resolution Land Characterization (MRLC) Consortium.

The MRLC Consortium is a partnership of federal agencies that produce or use land cover data. Partners include the USGS (National Mapping, Biological Resources, and Water Resources Divisions), USEPA, the U.S. Forest Service, and the National Oceanic and Atmospheric Administration.

The California NLCD set was produced as part of a project area encompassing portions of Federal Region 9, including the states of California, Nevada, and Arizona. This data set was produced under the direction of the MRLC Regional Land Cover Characterization Project of the USGS EROS Data Center (EDC), Sioux Falls, SD.

For further information, see the

USGS Land Cover Characterization Program web site at: <http://edcwww.cr.usgs.gov/programs/lccp/>

NOTE: This metadata document has been modified from the original USGS document to reflect data enhancements performed by CA Department of Transportation. Specific enhancement categories are indicated by “(Caltrans)”.

Projection Information (Caltrans):

VITAL STATISTICS:

Datum:	NAD 83
Projection:	Albers Conic Equal Area
Units:	Meters
Source:	U. S. Geological Survey
Data Scale:	1:100,000
Data Resolution	30 meter
1st Std. Parallel:	34 degrees N
2nd Std. Parallel:	40.5 degrees N
Longitude of Origin:	-120 degrees W
Latitude of Origin:	0 degrees N
False Easting:	0
False Northing:	-4,000,000
Source Projection:	(Cited below)
Source Units:	(Cited below)
Source Scale:	(Cited below)
Conversion Software:	ArcInfo v. 8.0.2
Data Structure:	GRID cell-based
ArcInfo Coverage Type:	Raster
ArcInfo Precision:	Double
ArcInfo Tolerances:	N/A

Naming Conventions (Caltrans):

The table below shows the 1x1 degree blocks necessary for displaying the Caltrans District extents:

DISTRICT	1	2	3	4	5	6
BLOCKS	D4112 4	D41123	D40121	D38123	D37122	D37118
	D4112 3	D41122	D39122	D38122	D37121	D37120
	D4012 4	D41121	D39121	D38121	D36122	D37119
	D4012 3	D41120	D39120	D37123	D36121	D36118
	D3912 4	D40123	D38122	D37122	D36120	D36120
	D3912 3	D40122	D38121	D37121	D35121	D36119
	D3912 2	D40121	D38120	D36121	D35120	D35117
	D3812 3	D40120	D38119		D35119	D35118
	D3812 2	D39123			D34120	D35120
		D39122			D34119	D35119
		D39121				D34117
		D39120				D34118
						D34119

DISTRICT	7	8	9	10	11	12
BLOCKS	D3411 7	D35114	D38118	D38121	D33114	D33117
	D3411 8	D35115	D38119	D38120	D33115	D33118
	D3411 9	D35116	D37117	D38119	D33116	
	D3311 7	D35117	D37118	D37121	D33117	
	D3311 8	D34114	D37119	D37120	D32114	
		D34115	D36115	D37119	D32115	
		D34116	D36116	D36121	D32116	
		D34117	D36117	D36120	D32117	
		D33114	D36118			
		D33115	D35115			
		D33116	D35116			
		D33117	D35117			
			D35118			

Data Display Notes (Caltrans – dated March 7, 2001):

There are two main options for displaying the data using the (USGS-modified) Anderson Land Cover Class Definitions:

For ArcInfo and ArcView users (excluding the Spatial Analyst extension): use the colormap file named “DXXXXX.clr”. Rename the “DXXXXX.clr” file to the GRIDs’ name (e.g. “D36117.clr”), then add the GRID theme as a “Image Data Source”.

Option for ArcView Spatial Analyst extension users: rename the “DXXXXX.avl” file to the GRID’s name (e.g. “D36117.avl”); then add the GRID as a “GRID data source”, and load the .avl file via the Legend Editor.

For information on modified Anderson land-use and land-cover classification system:

<http://edcwww.cr.usgs.gov/programs/lccp/classes.html>

Data Processing Notes (Caltrans – dated March 7, 2001):

The NLCD data was received as two Band Interleaved by Line (BIL) images, for the northern and southern halves of California. After performing coordinate shifts improve overlay correlation, the BIL images were converted and merged into a statewide ArcInfo GRID. Then the statewide GRID was reprojected into the Albers NAD83 parameters used by the Caltrans HQ GIS Data Library.

The initial Landsat TM mosaics, all ancillary data sets, and the land cover product are all map-registered to an Albers Conical Equal Area projection. The following represents the projection information (USGS) for the final land cover product for California.

Projection:	Albers	Conical Equal Area
Datum:		NAD83
Spheroid:		GRS80
Standard Parallels:		29.5 degrees North Latitude 45.5 degrees North Latitude
Central Meridian:		96 degrees West Longitude
Origin of the Projection:		23 degrees North Latitude
False Easting:		0 meters
False Northing:		0 meters
Number of Lines (rows/height):		20260
Number of Samples (columns/width):		14971
Number of Bands:		1 Pixel size: 30 X 30 meters
Projection Coordinates		(center of pixel, projection meters)
Upper Left Corner:		-2361900 meters(X), 2455830 meters(Y)
Lower Right Corner:		-1912800 meters(X), 1848060 meters(Y)

NOTE: Each state data set was extracted from the larger regional data set. State boundaries from the USGS 1:100,000 Digital Line Graph (DLG) series were used as the basis for extracting the state data. In many instances, the precision of the boundaries in the 1:100,000 DLG data does not match the spatial precision of the Landsat TM data. This is most apparent where state boundaries follow small rivers. To overcome the possibility of data being lost in the extraction process, a 300 meter (10 pixel) buffer was added to the state boundary used to extract the state data.

Version 1999-12:

----- Initial release of preliminary dataset.

Version 2000-04 Edits:

A number of zero data value pixels that were found in the preliminary dataset were corrected and some edgematching was performed.

Data Sources:

The base data set for this project was leaves-off Landsat TM data, nominal-1992 acquisitions. Other ancillary data layers included leaves-on TM, USGS 3-arc second Digital Terrain Elevation Data (DTED) and derived slope, aspect and shaded relief, Bureau of the Census population and housing density data, USGS land use and land cover (LUDA), and National Wetlands Inventory (NWI) data if available.

Landsat thematic mapper (TM) scenes:

Path	Row	Date
041	034	20 Apr 87
042	033	10 May 92
042	034	11 Jun 92
043	032	26 Apr 90
043	033	02 Jul 91
043	034	12 May 90
044	031	24 May 92
044	032	17 Jun 89
044	033	08 May 92
044	034	06 Apr 92
045	031	31 May 92
045	032	17 Mar 88
045	033	27 Apr 91
046	031	05 Jun 91
046	032	05 Jun 91

LEAF ON:

041	034	10 Aug 93
042	033	01 Aug 93
042	034	28 Aug 91
043	032	08 Aug 93
043	033	24 Aug 93
043	034	06 Sep 92
044	031	10 Aug 91
044	032	26 Aug 91
044	033	26 Aug 91
044	034	08 Sep 90
045	031	19 Aug 92
045	032	19 Aug 92
045	033	22 Aug 93
046	031	29 Aug 93
046	032	08 Oct 90

038	036	10 Apr 91
038	037	19 Mar 92
039	035	11 Apr 92
039	036	19 Apr 92
039	037	11 Apr 92
040	035	07 May 90
040	036	02 Apr 89
040	037	26 Apr 92
041	035	20 Apr 87
041	036	20 Apr 87
041	037	06 Jul 92
042	035	02 May 89
042	036	05 Apr 88
043	035	07 Apr 89
043	036	06 Sep 92
044	035	13 Sep 92

LEAF ON:

038	036	16 Aug 91
038	037	29 Aug 90
039	035	27 Jul 93
039	036	27 Jul 93
039	037	27 Jul 93
040	035	19 Aug 93
040	036	19 Aug 93
040	037	16 Aug 92
041	035	10 Aug 93
041	036	26 Aug 93
041	037	06 Jul 92
042	035	17 Aug 93
042	036	04 Sep 92
043	035	06 Sep 92
043	036	06 Sep 92
044	035	13 Sep 92

Caveats and Concerns:

1) Both irrigated and dryland agriculture are practiced in this region. In the dryland areas small grains predominate; fields are classified as fallow when there is no evidence of visible vegetation indicating a prescribed alternation between cropping and tillage. Crop types in the irrigated areas were difficult to reliably distinguish; row crops are likely to be under represented where no field observations or other ancillary information was incorporated.

Accuracy Assessment:

Accuracy assessment is pending.

21-Class National Land Cover Data Key:

NOTE - All Classes May NOT Be Represented in a specific state data set.
The class number represents the digital value of the class in the data set.

NLCD Land Cover Classification System Key - Rev. July 20, 1999

Water

- 11 Open Water
- 12 Perennial Ice/Snow

Developed

- 21 Low Intensity Residential
- 22 High Intensity Residential
- 23 Commercial/Industrial/Transportation

Barren

- 31 Bare Rock/Sand/Clay
- 32 Quarries/Strip Mines/Gravel Pits
- 33 Transitional

Forested Upland

- 41 Deciduous Forest
- 42 Evergreen Forest
- 43 Mixed Forest

Shrubland

- 51 Shrubland

Non-natural Woody

- 61 Orchards/Vineyards/Other

Herbaceous Upland

- 71 Grasslands/Herbaceous

Herbaceous Planted/Cultivated

- 81 Pasture/Hay
- 82 Row Crops
- 83 Small Grains
- 84 Fallow
- 85 Urban/Recreational Grasses

Wetlands

- 91 Woody Wetlands
- 92 Emergent Herbaceous Wetlands

NLCD Land Cover Classification System Land Cover Class Definitions

Water - All areas of open water or permanent ice/snow cover.

11. Open Water - All areas of open water; typically 25 percent or greater cover of water (per pixel).

12. Perennial Ice/Snow - All areas characterized by year-long cover of ice and/or snow.

Developed - Areas characterized by a high percentage (30 percent or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc).

21. Low Intensity Residential - Includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30-80 percent of the cover. Vegetation may account for 20 to 70 percent of the cover. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.

22. High Intensity Residential - Includes highly developed areas where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20 percent of the cover. Constructed materials account for 80 to 100 percent of the cover.

23. Commercial/Industrial/Transportation - Includes infrastructure (e.g. roads, railroads, etc.) and all highly developed areas not classified as High Intensity Residential.

Barren - Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen cover may be extensive.

31. Bare Rock/Sand/Clay - Perennially barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, beaches, and other accumulations of earthen material.

32. Quarries/Strip Mines/Gravel Pits - Areas of extractive mining activities with significant surface expression.

33. Transitional - Areas of sparse vegetative cover (less than 25 percent of cover) that are dynamically changing from one land cover to another, often because of land use activities. Examples include forest clearcuts, a transition phase between forest and agricultural land, the temporary clearing of vegetation, and changes due to natural causes (e.g. fire, flood, etc.).

Forested Upland - Areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100 percent of the cover.

41. Deciduous Forest - Areas dominated by trees where 75 percent or more of the tree species shed foliage simultaneously in response to seasonal change.

42. Evergreen Forest - Areas dominated by trees where 75 percent or more of the tree species maintain their leaves all year. Canopy is never without green foliage.

43. Mixed Forest - Areas dominated by trees where neither deciduous nor evergreen species represent more than 75 percent of the cover present.

Shrubland - Areas characterized by natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included.

51. Shrubland - Areas dominated by shrubs; shrub canopy accounts for 25-100 percent of the cover. Shrub cover is generally greater than 25 percent when tree cover is less than 25 percent. Shrub cover may be less than 25 percent in cases when the cover of other life forms (e.g. herbaceous or tree) is less than 25 percent and shrubs cover exceeds the cover of the other life forms.

Non-natural Woody - Areas dominated by non-natural woody vegetation; non-natural woody vegetative canopy accounts for 25-100 percent of the cover. The non-natural woody classification is subject to the availability of sufficient ancillary data to differentiate non-natural woody vegetation from natural woody vegetation.

61. Orchards/Vineyards/Other - Orchards, vineyards, and other areas planted or maintained for the production of fruits, nuts, berries, or ornamentals.

Herbaceous Upland - Upland areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100 percent of the cover.

71. Grasslands/Herbaceous - Areas dominated by upland grasses and forbs. In rare cases, herbaceous cover is less than 25 percent, but exceeds the combined cover of the woody species present. These areas are not subject to intensive management, but they are often utilized for grazing.

Planted/Cultivated - Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100 percent of the cover.

81. Pasture/Hay - Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops.

82. Row Crops - Areas used for the production of crops, such as corn, soybeans, vegetables, tobacco, and cotton.

83. Small Grains - Areas used for the production of graminoid crops such as wheat, barley, oats, and rice.

84. Fallow - Areas used for the production of crops that are temporarily barren or with sparse vegetative cover as a result of being tilled in a management practice that incorporates prescribed alternation between cropping and tillage.

85. Urban/Recreational Grasses - Vegetation (primarily grasses) planted in developed settings for recreation, erosion control, or aesthetic purposes. Examples include parks, lawns, golf courses, airport grasses, and industrial site grasses.

Wetlands - Areas where the soil or substrate is periodically saturated with or covered with water as defined by Cowardin et al.

91. Woody Wetlands - Areas where forest or shrubland vegetation accounts for 25-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

92. Emergent Herbaceous Wetlands - Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

General Procedures

Land Cover Characterization:

The project is being carried out on the basis of 10 Federal Regions that make up the conterminous United States; each region is comprised of multiple states; each region is processed in subregional units that are limited to the area covered by no more than 18 Landsat TM scenes. The general NLCD procedure is to: (1) mosaic subregional TM scenes and classify them using an unsupervised clustering algorithm, (2) interpret and label the clusters/classes using aerial photographs as reference data, (3) resolve the labeling of confused clusters/classes using the appropriate ancillary data source(s), and (4) incorporate land cover information from other data sets and perform manual edits to augment and refine the "basic" classification developed above.

Two seasonally distinct TM mosaics are produced, a leaves-on version (summer) and a leaves-off (spring/fall) version. TM bands 3 4 5 and 7 are mosaicked for both the leaves-on and leaves-off versions. For mosaicking purposes, a base scene is selected for each mosaic and the other scenes are adjusted to mimic spectral properties of the base scene using histogram matching in regions of spatial overlap.

Following mosaicking, either the leaves-off version or leaves-on version is selected to be the "base" for the land cover mapping process. The 4 TM bands of the "base" mosaic are clustered to produce a single 100-class image using an unsupervised clustering algorithm. Each of the spectrally distinct clusters/classes is then assigned to one or more Anderson level 1 and 2 land cover classes using National High Altitude Photography program (NHAP) and National Aerial Photography program (NAPP) aerial photographs as a reference. Almost invariably, individual spectral clusters/classes are confused between two or more land cover classes.

Separation of the confused spectral clusters/classes into appropriate NLCD class is accomplished using ancillary data layers. Standard ancillary data layers include: the "non-base" mosaic TM bands and 100-class cluster image; derived TM normalized vegetation index (NDVI), various TM band ratios, TM date bands; 3-arc second Digital Terrain Elevation Data (DTED) and derived slope, aspect and shaded relief; population and housing density data; USGS land use and land cover (LUDA); and National Wetlands Inventory (NWI) data if available. Other ancillary data sources may include soils data, unique state or regional land cover data sets, or data from other federal programs such as the National Gap Analysis Program (GAP) of the USGS Biological Resources Division (BRD). For a given confused spectral cluster/class, digital values of the various ancillary data layers are compared to determine: (1) which data layers are the most effective for splitting the confused cluster/class into the appropriate NLCD class, and (2) the appropriate layer thresholds for making the split(s). Models are then developed using one to several ancillary data layers to split the confused cluster/class into the NLCD class. For example, a population density threshold is used to separate high-intensity residential areas from commercial/industrial/transportation. Or a cluster/class might be confused between row crop and grasslands. To split this particular cluster/class, a TM NDVI threshold might be identified and used with an elevation threshold in a class-splitting model to make the appropriate NLCD class assignments. A purely spectral example is using the temporally opposite TM layers to discriminate confused cluster/classes such as hay pasture vs. row crops and deciduous forests vs. evergreen forests; simple thresholds that contrast the seasonal differences in vegetation between leaves-on vs. leaves-off.

Not all cluster/class confusion can be successfully modeled out. Certain classes such as urban/recreational grasses or quarries/strip mines/gravel pits that are not spectrally unique require manual editing. These class features are typically visually identified and then reclassified using on-screen digitizing and recoding. Other classes such as wetlands require the use of specific data sets such as NWI to provide the most accurate classification. Areas lacking NWI data are typically subset out and modeling is used to estimate wetlands in these localized areas. The final NLCD product results from the classification (interpretation and labeling) of the 100-class "base" cluster mosaic using both automated and manual processes, incorporating both spectral and conditional data layers. For a more detailed explanation please see Vogelmann et al. 1998 and Vogelmann et al. 1998.

Accuracy Assessment:

An accuracy assessment is done on all NLCD on a Federal Region basis following a revision cycle that incorporates feedback from MRLC Consortium partners and affiliated users. The accuracy assessments are conducted by private sector vendors under contract to the USEPA. A protocol has been established by the USGS and USEPA that incorporates a two-stage, geographically stratified cluster sampling plan (Zhu et al., 1999) utilizing National Aerial Photography Program (NAPP) photographs as the sampling frame and the basic sampling unit. In this design a NAPP photograph is defined as a 1st stage or primary sampling unit (PSU), and a sampled pixel within each PSU is treated as a 2nd stage or secondary sampling unit (SSU).

PSU's are selected from a sampling grid based on NAPP flight-lines and photo centers, each grid cell measures 15' X 15' (minutes of latitude/longitude) and consists of 32 NHAP photographs. A geographically stratified random sampling is performed with 1 NAPP photo being randomly selected from each cell (geographic strata), if a sampled photo falls outside of the regional boundary it is not used. Second stage sampling is accomplished by selecting SSU's (pixels) within each PSU (NAPP photo) to provide the actual locations for the reference land cover classification.

The SSU's are manually interpreted and misclassification errors are estimated and described using a traditional error matrix as well as a number of other important measures including the overall proportion of pixels correctly classified, user's and producer's accuracies, and omission and commission error probabilities.

Discussion:

While we believe that the approach taken has yielded a very good general land cover classification product for a large region, it is important to indicate to the user where there might be some potential problems. The biggest concerns are listed below:

- 1) Some of the TM data sets are not temporally ideal. Leaves-off data sets are heavily relied upon for discriminating between hay/pasture and row crop, and also for discriminating between forest classes. The success of discriminating between these classes using leaves-off data sets hinges on the time of data acquisition. When hay/pasture areas are non-green, they are not easily distinguishable from other agricultural areas using remotely sensed data. However, there is a temporal window during which hay and pasture areas green up before most other vegetation (excluding evergreens, which have different spectral properties); during this window these areas are easily distinguishable from other crop areas. The discrimination between hay/pasture and deciduous forest is likewise optimized by selecting data in a temporal window where deciduous vegetation has yet to leaf out. It is difficult to acquire a single-date of imagery (leaves-on or leaves-off) that adequately differentiates between both deciduous/hay and pasture and hay-pasture/row crop.

- 2) The data sets used cover a range of years (see data sources), and changes that have taken place across the landscape over the time period may not have been captured. While this is not viewed as a major problem for most classes, it is possible that some land cover features change more rapidly than might be expected (e.g. hay one year, row crop the next).
- 3) Wetlands classes are extremely difficult to extract from Landsat TM spectral information alone. The use of ancillary information such as National Wetlands Inventory (NWI) data is highly desirable. We relied on GAP, LUDA, or proximity to streams and rivers as well as spectral data to delineate wetlands in areas without NWI data.
- 4) Separation of natural grass and shrub is problematic. Areas observed on the ground to be shrub or grass are not always distinguishable spectrally. Likewise, there was often disagreement between LUDA and GAP on these classes.

Acknowledgments

This work was performed by the Raytheon STX Corporation under U.S. Geological Survey Contract 1434-CR-97-CN-40274.

References

More detailed information on the methodologies and techniques employed in this work can be found in the following:

Kelly, P.M., and White, J.M., 1993. Preprocessing remotely sensed data for efficient analysis and classification, Applications of Artificial Intelligence 1993: Knowledge-Based Systems in Aerospace and Industry, Proceeding of SPIE, 1993, 24-30.

Cowardin, L.M., V. Carter, F.C. Golet, and E.T. LaRoe, 1979. Classification of Wetlands and Deepwater Habitats of the United States, Fish and Wildlife Service, U.S. Department of the Interior, Oregon, D.C.

Vogelmann, J.E., Sohl, T., and Howard, S.M., 1998. "Regional Characterization of Land Cover Using Multiple Sources of Data." Photogrammetric Engineering & Remote Sensing, Vol. 64, No. 1, pp. 45-47.

Vogelmann, J.E., Sohl, T., Campbell, P.V., and Shaw, D.M., 1998. "Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data Sources." Environmental Monitoring and Assessment, Vol. 51, pp. 415-428.

Zhu, Z., Yang, L., Stehman, S., and Czaplewski, R., 1999. "Designing an Accuracy Assessment for USGS Regional Land Cover Mapping Program." (In review) Photogrammetric Engineering & Remote Sensing.